

同一名詞の反復から見たジャンル間の文体差とその 要因 - コーパスを用いた定量的分析を通して -

著者	鯨井 綾希
雑誌名	言語科学論集
巻	16
ページ	13-25
発行年	2012-12-01
URL	http://hdl.handle.net/10097/55289

同一名詞の反復から見たジャンル間の文体差とその要因 ーコーパスを用いた定量的分析を通してー

鯨 井 綾 希

キーワード: 同一名詞の反復、文体差、定量的分析、CSJ、BCCWJ

要 旨

『日本語話し言葉コーパス』と『現代日本語書き言葉均衡コーパス』を用い、ジャンルごとの同一名詞の反復(反復語)の様相を明らかにした。また、反復語の様相と個々のジャンルの文体的特徴との関係を分析した。考察の結果、話し言葉的な表現ほど反復の距離が短くなること、書き言葉的な表現ほど反復語の使用にばらつきが生じること、反復語の使用率や一語あたりの反復頻度には改まりの程度や文字数の制約が関わることが分かった。

1. はじめに

本稿では、『日本語話し言葉コーパス』(以下CSJ)と『現代日本語書き言葉均衡コーパス』(以下BCCWJ)を利用して、文章・談話のまとまりや意味的連続性を形成するのに重要な役割を担う同一名詞の反復(以下これを反復語と記述)の使用実態を定量的な側面から明らかにし、その上で話し言葉と書き言葉に見られる種々の表現上の特性を考慮して、反復語の量的側面から見たジャンル間の文体的な異同とその要因について分析することを目的とする。

2. 調査上の対象・資料・方法

2-1. 調査対象

本稿で対象とするのは、以下の波線部で示したような文章・談話内における内容を表す名詞¹の反復使用(以下これを反復語と表記)である。

- (1) 産業技術が極めて幼稚なうちは、思い込みで適当なことをやっても一見通用するように見える時代がある。しかし、どのような分野でも 産業技術は次第に高度化する。その結果、学問に基づいた本物の 技術しか通用しないということが

いろいろな産業分野でいま始まっている。

(大見忠弘2004『復活!日本の半導体産業-未来を拓く志-実力を磨いて世の中の役に立とう!-』財界研究所、『BCCWJ』サンプルID:PB45_00024)

(2) R: (F ええ)で猿だったら(F えーとー)聞こえんのか聞こえないのかっていうので(F えーっとですねー)(F えー)何だろう例えばピアノの音とかだと

L: (F はい)

R: (F えー)色んな音の成分がだくさん入ってるんですね

L: (F はい)

R: んでその(D い)一番低い音の成分から(F そのー)高い音の成分まで色々入ってるんだけど

(『CSJ』ID: D04M0056を見やすく整形)

(1)も(2)も、「産業」「技術」「音」という名詞が反復的に使用されることでそれを中心とした内容上の展開が構築されている。このような反復語の使用は、文章・談話中における文の意味連鎖やコミュニケーション上の機能に関わるものとして、従来から話し言葉・書き言葉の両方で研究が行われてきた(中田1991、田中1997、塩澤2005、馬場2006など)。しかし、反復語が文章・談話内でどの程度現れ、それがどのようなテキスト上の特徴と関係しているのかという側面を定量的に分析した研究は、山崎(2010、2012)や鯨井(2012)などに関連するものが見られるものの多いとは言えず、その実態はそれほど詳細に把握されていないと考えられる。したがって、本稿ではコーパスを利用して様々なジャンルにおける反復語の様相を定量的に分析し、個々のジャンルによって反復語の使用にどのような量的差異が見られるかを明らかにする。また、そこで見られた差異をジャンルごとの特徴と照らし合わせ、その諸要因について論じる。

2-2. 調査資料

扱う資料は、CSJおよびBCCWJである。CSJは話し言葉のデータであり、本稿では対話を収録したデータと、学会講演と模擬講演という二つの独話データの三つを取り上げる。対話データには講演のインタビュー・課題指向対話・自由対話が含まれるが、それぞれのファイル数の少なさから、一括して扱う。

CSJは、ジャンルとしては話し言葉による音声的伝達という特徴を共通に持つ

コーパスである。しかし、対話は話し手・聞き手が随時交替可能なものであるが、講演のような独話は固定された話し手・聞き手によって成立するという点で、対話と講演は話し言葉としての様相が異なる。また、学会講演と模擬講演は、学会講演が改まった場所として設定されている一方、模擬講演は必ずしもそのような設定が存在していない点で、やはりジャンルとして異質であると考えられる。したがって、本稿ではCSJを話し言葉という大きな枠組みによるジャンルとしてとらえつつ、さらに対話・学会講演・模擬講演という下位のジャンルを設定する。

また、話し言葉という大きなジャンルと対比される存在として書き言葉がある。本稿では、書き言葉としてBCCWJを取り上げるが、書き言葉でも、そのジャンルとしてのあり方は多様である。本稿では、BCCWJに含まれる白書・新聞・書籍・雑誌をジャンルとして取り上げる。なお、BCCWJは正確な付加情報が与えられているコアデータのみ取り上げる²。調査資料の大きさは以下の表1に示す³。

表 1：分析資料の概略

	白書		新聞		書籍		雑誌	
	延べ語数	異なり語数	延べ語数	異なり語数	延べ語数	異なり語数	延べ語数	異なり語数
最小値	216	123	127	97	126	64	134	102
第1四分位値	701.5	238	230	144	254	144.5	273	177.8
中央値	981.5	303	275	169	438	215	521.5	291
平均値	1182.1	339.1	305.1	184.7	548.5	262	598.8	307.3
第3四分位値	1594.2	441.2	342.2	206.2	722	367.5	834.5	381.5
最大値	2925	713	1203	571	1593	719	1954	876
ファイル数	62		340		83		86	
	学会講演		模擬講演		対話			
	延べ語数	異なり語数	延べ語数	異なり語数	延べ語数	異なり語数		
最小値	119	82	193	97	176	82		
第1四分位値	590	175	278.5	139.5	291.5	119.2		
中央値	717	207	337	164	342	146.5		
平均値	778	226.1	345	168.1	365.1	151.2		
第3四分位値	882	256	403	195	404	171.8		
最大値	3509	917	659	280	922	268		
ファイル数	987		107		58			

2-3. 調査方法

定量的分析に際しては、名詞50語あたりの反復語の平均使用率と使用率のばらつき、反復語1語あたりの平均使用頻度と平均反復間隔という四つの側面に注目し、そ

れらをファイルごとに計算することで各ジャンルにおける計算結果の記述を行う。その上で各ジャンルの性質を考慮に加え、計量結果と対照させることで、文章・談話における反復語の現れ方を、ジャンルの性質との関係から明らかにしていく。なお、分析に際しては、「語」相当の集計単位として、CSJとBCCWJに付与されている形態論情報である「短単位」を利用した⁴。

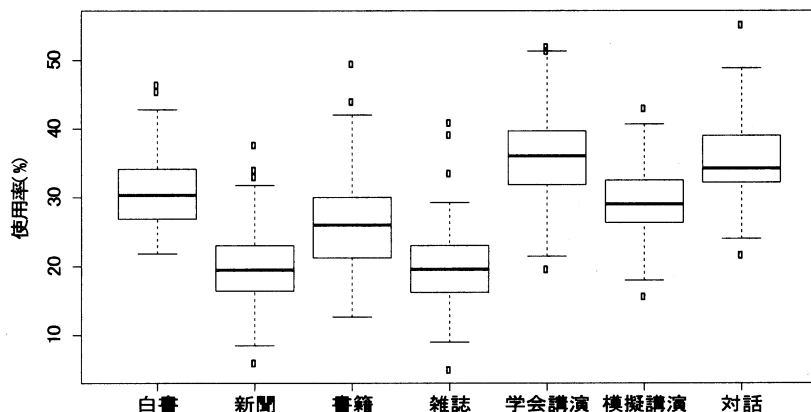
3. 調査結果

3-1. 平均使用率

反復語が文章・談話中でどの程度現れるのかを明らかにするために、始めに名詞50語ごとの反復語の平均使用率を計算した。反復語の使用率は、延べ語数に対する2回目以降に使用された名詞の割合によって求めた。

50語ごとに計算したのは、文章の長さに影響を受けない形で対象ファイルにおける反復語の使用率を算出するためである⁵。この方法に基づいた、対象資料ごとの各ファイルにおける反復語の平均使用率と、その箱ひげ図を図表1に示す⁶。

	白書	新聞	書籍	雑誌	学会講演	模擬講演	対話
最小値	21.81	6.00	12.67	5.00	19.62	15.67	21.67
第1四分位値	26.90	16.48	21.23	16.34	31.85	26.33	32.22
中央値	30.31	19.50	26.00	19.59	36.00	29.00	34.20
平均値	31.20	19.80	25.56	20.04	35.78	29.42	35.28
第3四分位値	34.07	23.00	30.00	23.00	39.66	32.50	38.96
最大値	46.46	37.67	49.50	40.91	52.00	43.00	55.20



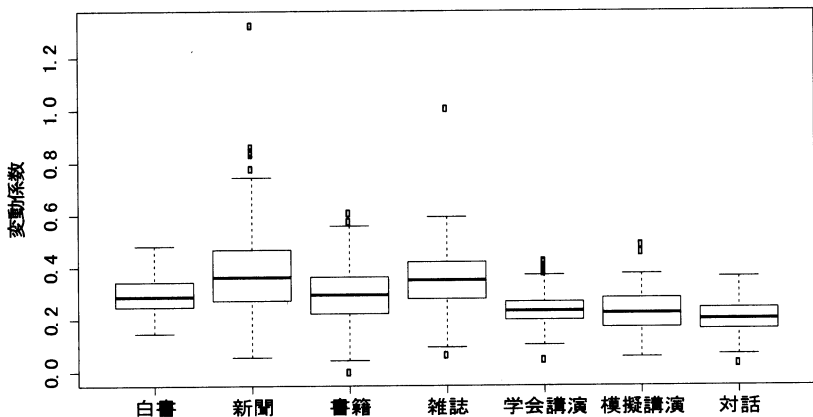
図表1：平均使用率(%)

50語あたりの反復語の使用率は、書き言葉に比べて話し言葉の方が総じて高い値を示す。また、書き言葉内では、白書の方が書籍よりも使用率が高く、書籍は雑誌や新聞よりも使用率が高い。また、話し言葉のうち、学会講演の中央値が模擬講演や対話に比べてやや高い値を取る。

3-2. 50語ごとの使用率のばらつき

前節では各ファイルで用いられる名詞50語ごとの反復語の平均使用率を計ったが、50語単位でくり返し計測しているため、一つのファイル内であっても計算ごとの結果にはばらつきが生じる。そこで、各ファイルで、50語ごとの使用率にどの程度のばらつきが見られるのか調査した結果を図表2に示す。なお、本稿のばらつきは変動係数によって表した。

	白書	新聞	書籍	雑誌	学会講演	模擬講演	対話
最小値	0.147	0.055	0.000	0.065	0.046	0.056	0.032
第1四分位値	0.248	0.272	0.222	0.280	0.198	0.167	0.163
中央値	0.287	0.362	0.295	0.350	0.232	0.224	0.200
平均値	0.293	0.381	0.293	0.358	0.234	0.230	0.203
第3四分位値	0.343	0.467	0.362	0.417	0.268	0.282	0.243
最大値	0.481	1.324	0.609	1.006	0.424	0.486	0.363



図表2：平均使用率のばらつき方

書き言葉のうち、白書を除くBCCWJの各ジャンルは最大値と最小値の差が大きく、コーパス内のファイルごとの反復語の使用率のばらつきそのものが大きい。つま

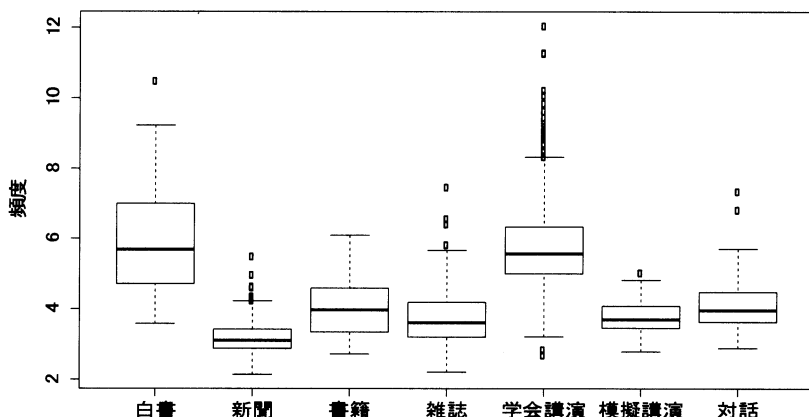
り、書き言葉には反復語の使用・不使用に関する一貫性が必ずしも見られず、量的に多様な様相を見せると考えられる。

また、変動係数の値の幅は、話し言葉の方が書き言葉よりも総じて小さい。したがって、話し言葉の方が場面にかかわらず反復語を用い、書き言葉の方が反復語を使ったり使わなかったりする傾向にあることが分かる。

3-3. 1語あたりの平均反復頻度

反復語は2回目以降使用された語が全て含まれるため、反復語に含まれる見出し語自体は多様である。したがって、反復語の使用実態を記述するために、その一つひとつの語がどの程度反復されているのかを把握することも有意義であると考えられる。各ファイルの反復語に含まれる見出し語1語あたりの平均頻度を集計した結果が以下の図表3である。

	白書	新聞	書籍	雑誌	学会講演	模擬講演	対話
最小値	3.58	2.14	2.72	2.21	2.70	2.79	2.89
第1四分位値	4.76	2.88	3.35	3.21	5.00	3.46	3.64
中央値	5.69	3.11	3.97	3.61	5.57	3.70	3.96
平均値	5.90	3.18	4.04	3.77	5.74	3.77	4.17
第3四分位値	6.98	3.42	4.59	4.18	6.34	4.09	4.48
最大値	10.49	5.50	6.09	7.47	12.06	5.03	7.36



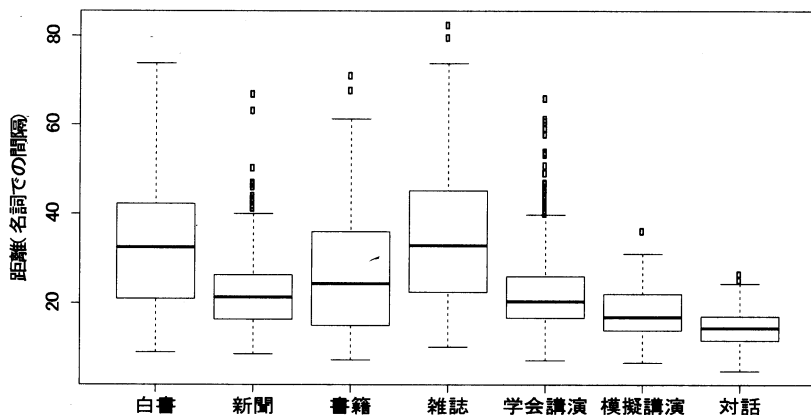
図表3：反復語1語あたりの平均反復頻度

図表3によれば、全体としては白書と学会講演の二つにおいて反復語となる見出し語1語あたりの使用頻度が高いことを指摘できる。それに比べ、図表1や2で見られたような、話し言葉と書き言葉の差は小さい。4節で詳述するが、白書と学会講演は改まりの程度が高いという点で性質が類似しており、ここでの結果もそのジャンル上の性質が関係していると考えられる。つまり、反復語として用いられる見出し語1語あたりの頻度は、書き言葉・話し言葉の別なく、改まりの程度が高くなると増加してくると言える。また、新聞は値の幅、値そのものの両方において小さく、字数制限による反復語使用の抑制が伺える。

3-4. 1語あたりの平均反復間隔

反復語は文章・談話中で複数回用いられる語の集合である。そのため、反復語となった語は、それぞれがある間隔を置いて改めて使用されており、前出の語と再使用された語との間には距離が生まれる。反復語に位置づけられる語のそれぞれにおいて、どの程度の間隔で反復が行われているのかを調査し、その平均間隔を計算した結果が図表4ならびに図表5である。図表4では名詞をもとにして距離を計測し、図表5では品詞にかかわらず全ての語によって距離を計測した。始めに、名詞の語数を利用した結果である図表4を以下に示す。

	白書	新聞	書籍	雑誌	学会講演	模擬講演	対話
最小値	9.00	8.57	7.27	10.15	7.16	6.64	4.77
第1四分位値	21.08	16.40	15.01	22.79	16.70	13.89	11.64
中央値	32.53	21.35	24.39	32.91	20.42	16.90	14.44
平均値	32.83	22.43	28.09	35.33	22.17	18.00	14.77
第3四分位値	41.76	26.31	36.01	44.93	26.01	22.08	17.02
最大値	73.78	66.96	71.11	82.50	66.00	36.28	26.60



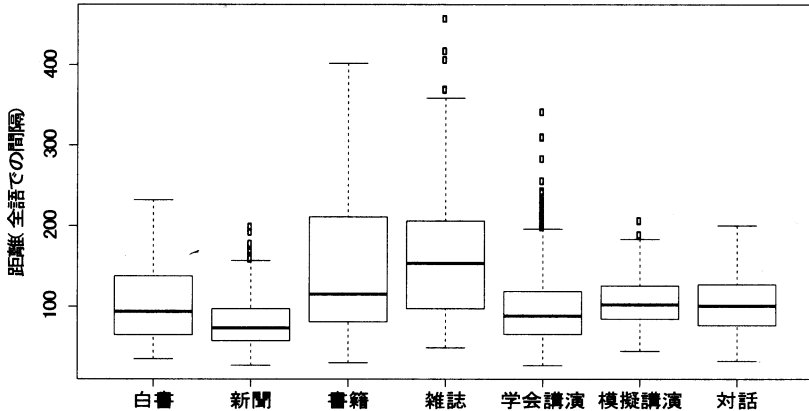
図表4：反復語1語あたりの平均反復間隔(名詞での間隔)

名詞間の測定では、BCCWJとCSJでは、CSJの方が明らかに間隔が短い。よって、書き言葉よりも話し言葉の方が短い間隔で語が反復されていると言える。また、話し言葉の中でも対話は最も反復の間隔が短くなる。

一方、書き言葉の中で見た場合、特に最大値において新聞が他の書き言葉と大きく異なった値を示している。したがって、ここでも字数制限による影響が考えられる。

次に、全ての語を利用して反復の間隔を測定したものが以下の図表5である。

	白書	新聞	書籍	雑誌	学会講演	模擬講演	対話
最小値	34.69	26.61	29.80	48.60	26.59	44.39	31.84
第1四分位値	65.23	57.36	80.92	98.08	65.50	84.42	76.76
中央値	93.62	73.23	115.32	153.57	88.31	102.45	100.85
平均値	101.37	80.28	147.44	169.45	96.58	108.40	102.98
第3四分位値	135.36	97.14	211.08	205.13	118.83	125.70	126.92
最大値	232.24	199.54	402.05	458.38	342.13	206.98	200.28



図表5：反復語 1語あたりの平均反復間隔(全語での間隔)

図表5では、雑誌を除く各ジャンルの中央値が概ね100語前後の値となっている。つまり、同一の名詞を反復させる間隔そのものは、話し言葉・書き言葉にかかわらず、基本的に100語前後が平均的であると考えられる。ただし、雑誌の中央値は150語程度となり、他に比べて間隔が大きい。また、全体としては、書籍や雑誌に見られるように書き言葉は最大値の値が大きくなる傾向にあり、反復語の使用間隔を大きくとることに対しての許容度が高いと考えられる。

新聞については、図表4と同様、書き言葉の中では例外的に最大値が小さくなっており、字数制限の影響が存在すると考えられる。また、白書は名詞のみに注目した場合の間隔においては書籍や雑誌と大差がなかったが、全ての語を測定に利用すると、書籍・雑誌に比べて最大間隔が小さくなる。

話し言葉においては、学会講演に外れ値が多くあるが、全体としては似通った値を示す。

4. ジャンルごとの性質から見た調査結果の解釈

以上の結果を解釈していくために、本節ではジャンルの性質について考察し、ジャンル間の関係をとらえなおす。CSJとBCCWJは、意思伝達の方法が音声である話し言葉か、書記である書き言葉かというジャンル上の大きな違いがある。また、複数の話し手が話す・聞くという関係を相互に行う点で、CSJにおける対話は書き言葉と決

定的に異なり、典型的な話し言葉として認められる。一方、白書のような公的機関による書き言葉は、話し言葉的要素をあまり導入しない点で書き言葉の典型と認めうる。

CSJの学会講演や模擬講演は一人の話し手によってなされ、筋書きが大方決められている点で書き言葉の音声化とも呼べる。したがって、この二つは対話に比べて書き言葉に近い存在である。また、話し言葉(音声言語)と書き言葉(書記言語)のそれぞれでも、改まりの程度に違いがあり、白書と学会講演は共に公的なものであるという点で改まりの程度が高い。書籍と模擬講演は、必ずしも改まりの場として捉えなくて良いという点で白書や学会講演よりも改まりの程度が低い。雑誌はそれらに比べるとさらに改まりの程度の低いものであると言える。新聞は改まりの程度としては白書同様の高さであると考えられるが、多くの情報を限られた字数内に収めなければならないという字数制限による表現上の制約が強く、その点で他のジャンルに対して特異な性質を持つ。

以上から、本稿で対象としたジャンルごとの関係は、音声言語・書記言語、独話型・対話型、改まり度の高低、字数制限という四つの基準からとらえられる。それらの基準とジャンルとの関係を図示したものが下の図1である⁷⁾。

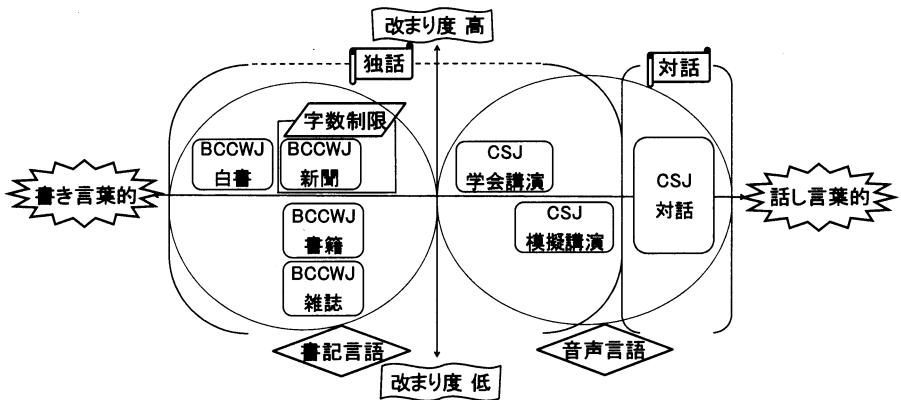


図1: ジャンルごとの差異と関係性

図1で示した枠組みによって、それぞれのジャンルにおける反復語の使用実態を、その文体的特徴との関係でとらえなおすことができる。反復語の使用実態から個々の文体的特徴を見た場合、話し言葉と書き言葉という大きな枠組みで最も解釈しや

すい文体差は平均使用間隔(図表4と5)であると考えられる。基本的には、平均使用間隔は書き言葉において大きな値を取ることが可能で、話し言葉的になるほど反復の間隔が短くなる。なお、平均使用間隔については、測定に用いる単位を全ての語に広げると(図表5)、話し言葉・書き言葉にかかわらず概ね100語程度が平均的な反復間隔となる。ただし、反復する語の種類や反復の意味が個々の語の使用において異なっていると考えられるため、100語という間隔それ自体にどのような意味があるのかという点については、現状では明確に述べることができない。この点は今後の課題とした。

「話し言葉的-書き言葉的」という観点からの文体差は、反復語の使用の一定性・ばらつき(図表2)についても当てはまると考えられる。この場合、話し言葉的になるほど、反復語を一貫して用いるようになり、書き言葉においては、反復語を使用したりしなかったりといったばらつきが大きくなる。

反復語の使用率(図表1)についても、全体としては話し言葉の方が書き言葉よりも使用率が上がっており、同様の対比の中で捉えられる。しかし、反復語の使用率では書き言葉で白書が、話し言葉で学会講演が高くなることから、そうした対比に加えて改まりの程度差も決定要因として関わっていると考えられ、典型的な話し言葉と書き言葉の対比構図の中だけでは捉えにくくなる。

反復語の見出し語1語あたりの平均頻度(図表3)においては、話し言葉的・書き言葉的という連続性はあまり大きな意味を持たない。白書と学会講演で頻度が高くなることから、改まりの程度によってその量を変動させていると言える。また、反復語の見出し語1語あたりの平均頻度(図表3)や平均使用間隔(図表4と5)では、書き言葉の中で新聞というジャンルが特異な値を示している。反復語の使用においては、1語あたりの頻度と間隔という点で字数制限が強い力を持って作用していると考えられる。

5. まとめ

本稿では、CSJとBCCWJに収録された様々なジャンルを利用し、反復語の量的側面から見たそれらの文体的な異同とその要因について分析することを目的とした。対象としたジャンルは話し言葉として対話・学会講演・模擬講演の三種類を、書き言葉として白書・新聞・書籍・雑誌の四種類を取り上げた。

各ジャンルの反復語の使用の様相とそれぞれのジャンルとしての性質を照らし合

わせると、両者の関係は次の表2のようにまとめることができる。

表2：反復語の使用実態に関わる諸要因

	音声言語-書記言語の影響	改まりの高低の影響	対話-独話の影響	字数制限の強弱の影響
反復語の使用率	有り 音声 \geq 書記	有り 高 \geq 低	無し ϕ	無し ϕ
反復語使用のばらつき	有り 音声 \leq 書記	無し ϕ	無し ϕ	無し ϕ
反復語1語あたりの頻度	無し ϕ	有り 高 \geq 低	無し ϕ	有り 強 \leq 弱
反復語の使用間隔	有り 音声 \leq 書記	無し ϕ	有り 対話 \leq 独話	有り 強 \leq 弱

また、この結果から、それぞれの文体の特徴とその関係性について、以下の点も指摘できる。すなわち、CSJとBCCWJという二つの資料において最も大きなジャンル上の差であると考えられる話し言葉と書き言葉という区別は、反復語の使用において必ずしも決定的要因となっているわけではないという点である。反復語の使用の様相は、その使用率や1語あたりの頻度、使用間隔において見られたように、改まりの程度や字数制限のような要因によっても変化すると考えられる。

なお、調査結果の違いはあくまでも量的差異によるものであるため、それが具体的にどのような質的違いの中で現れたものなのかについてはテキストの内実の観察による分析により達成されなければならない。この点は今後の課題としたい。

注

¹ 名詞のうち、具体的意味が希薄な「物(もの)・事(こと)・頃(ころ)・為(ため)・間(あいだ)・内(うち)・時(とき)・程(ほど)・由(よし)・後(あと)・項・まま・せい・はず・かた・ふし・ところ・ゆえ」と、数詞、ならびに名詞に位置づけられるが単独で用いられず関係節を構成して使われる「以上・以前・場合・辺り(あたり)」は除いた。

² コアデータにはその他にも Web 上のブログや知恵袋という質問サイトのデータも存在するが、データ内に存在する各ファイルが対象にできる大きさに至っていないため、本稿では扱わない。また、本稿で利用した BCCWJ は、2011 年 3 月に配布された「特定領域研究「日本語コーパス」研究成果報告 DVD」内に含まれるデータであるが、BCCWJ は 2012 年 9 月現在、完成版の DVD が公開されている。

³ ここで単位となる「語」の認定基準は後述。また、同じく後述するが、本稿では 50 語ごとの反復語の使用率とそのばらつきを分析に取り入れるため、延べ語数 100 語未満のファイルは分析対象から外した。

⁴ 短単位については小椋(2006)、小椋(2008)、小椋ほか(2011)を参照。なお、それを用いた集計や計算、作図はスクリプト言語 Perl と統計解析環境 R を利用している。

⁵ 50 語ごとに集計した際の余りは切り捨てた。

⁶ 箱ひげ図は、箱内の線が中央値であり、下辺と上辺がそれぞれ第 1 四分位値と第 3 四分位値を指す。上下の

丸は外れ値、上下の点線の先は外れ値を除いた場合の最大値と最小値である。

⁷ 図1で学会講演が模擬講演よりも書き言葉的であるとしているが、これは典型的な書き言葉である白書と改まりの程度において類似性が高いのが学会講演であると判断したことによる。

参考文献

- 井上次夫(2011)「書き言葉らしさの判断と測定」『特定領域研究「日本語コーパス」平成22年度公開ワークショップ(研究成果報告)予稿集』, pp. 89 - 96.
- 小椋秀樹(2006)「第3章形態論情報」『国立国語研究所報告124 日本語話し言葉コーパスの構築法』, pp. 133 - 186.
- 小椋秀樹(2008)「『日本語話し言葉コーパス』の言語単位」『日本語学』27. 5, pp. 72 - 81.
- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕(2011)「『現代日本語書き言葉均衡コーパス』形態論情報規定集第4版(上)(下)」国立国語研究所内部報告集LR - CCG - 10 - 05 - 01, LR - CCG - 10 - 05 - 02(特定領域研究「日本語コーパス」研究成果報告DVD所収)。
- 鯨井綾希(2012)「文章における名詞の反復の量的様相 - Type - Token Ratioを利用した分析 - 」『計量国語学』28 - 6, pp. 211 - 225.
- 小磯花絵・田中弥生・小木曾智信・近藤明日子(2011)「テキストの多様性をとらえる分類指標の構築を目指して」『特定領域研究「日本語コーパス」平成22年度公開ワークショップ(研究成果報告)予稿集』, pp. 431 - 442.
- 塩澤和子(2005)「コラムに観察されるくり返しの機能」『文藝言語研究言語篇』47, pp. 15 - 31.
- 田中妙子(1997)「会話におけるくりかえし」- テレビ番組を資料として - 、『早稲田大学日本語研究教育センター紀要』9, pp. 47 - 67.
- 中田智子(1991)「会話にあらわれるくり返しの発話」『日本語学』10. 10, pp. 52 - 62.
- 馬場俊臣(2006)『日本語の文連接表現 - 指示・接続詞・反復 - 』おうふう。
- 山崎誠(2010)「語の平均使用頻度に現れるテキストの特徴」『特定領域研究「日本語コーパス」平成21年度公開ワークショップ(研究成果報告会)予稿集』, pp. 5 - 14.
- 山崎誠(2012)「共起語率の分布からみるテキストの語彙的特徴」『第1回コーパス日本語学ワークショップ予稿集』, pp. 221 - 226.

引用・調査資料

- 「日本語話し言葉コーパス」(科学技術振興調整費開放的融合研究「話し言葉の言語的・パラ言語的構造の解明に基づく「話し言葉工学」の構築」)。
- 「現代日本語書き言葉均衡コーパス・コアデータ」(特定領域研究「日本語コーパス」データ班, 特定領域研究「日本語コーパス」研究成果報告DVD(JC - G - 10 - 04)所収)。

関連URL

- The Perl Programming Language(<http://www.perl.org/>) 2012/9/28閲覧。
- The R Project for Statistical Computing(<http://www.r-project.org/>) 2012/9/28閲覧。